Predicting pedestrian road-crossing assertiveness for autonomous vehicle control

F. Camara*†, O. Giles*, R. Madigan*, M. Rothmüller‡ P. Holm Rasmussen‡,
S. A. Vendelbo-Larsen‡, G. Markkula*, Y. M. Lee*, Laura Garach§, N. Merat*, C. W. Fox* † ¶
*Institute for Transport Studies, University of Leeds, UK
†Lincoln Center for Autonomous Systems, University of Lincoln, UK
‡ School of Architecture, Design and Planning, Aalborg University, DK
§ Department of Civil Engineering, University of Granada, ES
¶ Ibex Automation Ltd., UK

***Abstract*—Autonomous vehicles (AVs) must interact with other road users including pedestrians. Unlike passive environments, pedestrians are active agents having their own utilities and decisions, which must be inferred and predicted by AVs in order to control interactions with them and navigation around them. In particular, when a pedestrian wishes to cross the road in front of the vehicle at an unmarked crossing, the pedestrian and AV must compete for the space, which may be considered as a game-theoretic interaction in which one agent must yield to the other. To inform AV controllers in this setting, this study collects and analyses data from real-world human road crossings to determine what features of crossing behaviours are predictive about the level of assertiveness of pedestrians and of the eventual winner of the interactions. It presents the largest and most detailed data set of its kind known to us, and new methods to analyze and predict pedestrian-vehicle interactions based upon it. Pedestrian-vehicle interactions are decomposed into sequences of independent discrete events. We use probabilistic methods – logistic regression and decision tree regression – and sequence analysis to analyze sets and sub-sequences of actions used by both pedestrians and human drivers while crossing at an intersection, to find common patterns of behaviour and to predict the winner of each interaction. We report on the particular features found to be predictive and which can thus be integrated into game-theoretic AV controllers to inform real-time interactions.**

***Index Terms*—Human Factors, Agent-Human Interactions, Autonomous Vehicles**

I. Introduction

While localisation, mapping, route planning, and control are now largely solved problems for autonomous vehicles in static environments [10], the major outstanding challenge for real-world autonomous vehicles is to operate in environments containing people. Unlike static (and ballistic) environments, people are complex interactive agents having their own goals, utilities, and decision making systems, and interactions with them must take these into account in order to predict their actions and plan accordingly. Interaction is recursive and complex: the AVs own actions will affect the persons actions and vice versa over time. This is most critical in environments where traffic rules do not clearly define priority for any of the participants, such as at unmarked intersections, where AVs and pedestrians have to negotiate with one another for priority. Conflict rates at unsignalized intersections are much higher than in other types of intersections [12] because the priority is not defined, and each agent acts based on their own interpretation. A mathematical model of such interactions based on game theory was recently presented [7] which proves that (under several assumptions including discretisable space and time, no lateral motion and communication, only via agent positioning on the road, the optimal strategy for both agents is probabilistic and recursive. Under this strategy, as the two agents get closer over time, both should gradually increase the probability that they will yield at each time, then draw their yield or non-yield action from this probability. These probabilities tend to unity as the agents positions get closer to a collision occurring. But importantly, the model proves that there must remain some small but strictly non-zero probability of the crash actually occurring, in both agents strategies, in order for the interaction to proceed optimally. A second study [3], then empirically measured human behaviours in a laboratory version of a road crossing scenario, and showed that it is possible to assign a single parameter to each agent which summarizes their entire behavioral preferences during such interactions. This parameter measures "assertiveness" as $\theta = U_{time}/U_{crash}$, the ratio of the agents value of time (i.e. the dollar value of losing 1 second of arriving at their destination, for example by yielding to the other agent for road priority), and the agents (negative) value of the collision actually occurring (which will be worse for an unarmoured pedestrian than for the driver of a heavy protective car, especially of a larger car such as an SUV).

Controllers based on this game theoretic model would thus benefit from any additional information about θ. The previous study recommended future study to search for externally observable variables which may be predictors of θ. For example, it is a priori possible that fixed demographic factors such as age and gender will predict θ, and/or that interactions-specific events such as the presence or non presence of eye-contact or signals will correlate with θ.

The present study proposes a new method of analyzing pedestrian-vehicle interactions to this end. We cannot observe θ directly for human pedestrians but we can, as a proxy, observe the final outcome of interactions between them and (human driver) vehicles in the field, along with factors which may have predictive value to both the winner and thus to θ. We designed the most detailed and largest data collection exercise known to us to collect accurate manual annotated observations of 204 road-crossing interactions including the presence or absence of 62 individual temporal event types and 12 environmental descriptors within each interaction. We perform sequence analysis to discover and report common short n-gram motifs from these sequences, and use these as additional features for outcome prediction. Lastly, we apply two regression models – logistic and decision tree regressions – to discover which events and motifs are informative to predict the winner for each interaction, and which thus may also provide useful information about assertiveness θ.

This project has received funding from EU H2020 interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395.

A. Related work

To our knowledge, there is no previous work related to pedestrian-vehicle crossing behaviour dealing with motifs extraction and regression models to potentially determine the relevant sequences of actions useful for winning the interaction. A review on different approaches for pedestrian crossing behavior modelling and analysis is provided in [13]. Methods of analysis are often performed via video recording, semi-structured interviews and VR recording. Previous work on pedestrian crossing behavior analysis can be found in [18] [8] [12] [15] [25]. Rasouli et al. introduced [18] [19] a novel dataset composed of 650 video-clips for driver-pedestrian interactions in several locations and different weather conditions. The analysis of their data show that attention plays an important role, as in 90% of the time, pedestrians reveal their intention of crossing by looking at the approaching vehicles. Rasouli et al. also presented some behavioral patterns that have been observed in their data, that show some frequent sequences of actions that are used by pedestrians in their crossing behavior. Similar to our approach, [17] used task analysis to divide pedestrian-vehicle interaction as a sequence of actions giving two outcomes, either the vehicle passes first or the pedestrian crosses, and performed some experiments with participants on their crossing behavior using virtual reality. In [8], Gorrini et al. analyzed video data of interaction between pedestrians and vehicles at an unsignalized intersection using semi-automatic tracking. Their study shows that pedestrian crossing behaviour can be divided into 3 phases: approaching (stable speed), appraising (deceleration due to evaluation of speed and distance of oncoming vehicles) and crossing (acceleration). Papadimitriou et al. [15] made a comparison of observed and declared behaviour of pedestrians at different crossing areas, as a method to assess pedestrian risk taking while crossing. They found that pedestrians' observed behaviour is in accordance with their declared behaviours from a questionnaire survey and they report that female and male participants have similar crossing behaviour. In [12], drivers' crossing behaviour model in China at unsignaled intersections is presented using game theory and their risk perception is inferred via an adaptative neuro-fuzzy inference system. Previous works [23] [12] [14] have focused on the evaluation of speed, TTC (Time To Collision), gap acceptance and communication means (e.g eye contact and motion pattern) of the road users but not really into how the interaction can be modelled as a sequence of actions, more meaningful for autonomous systems.

Motif analysis, widely used in bioinformatics, consists in finding some biological patterns in the genomic sequence [6]. The DNA being composed of the four nucleotides A, T, C, G, a motif can be seen as a "short sequence". Motif prediction may be performed via supervised or unsupervised learning. The MEME algorithm (Multiple EM for Motif Elicitation) has been applied to genetic sequence analysis but also to various other domains such as musical audio analysis [1].

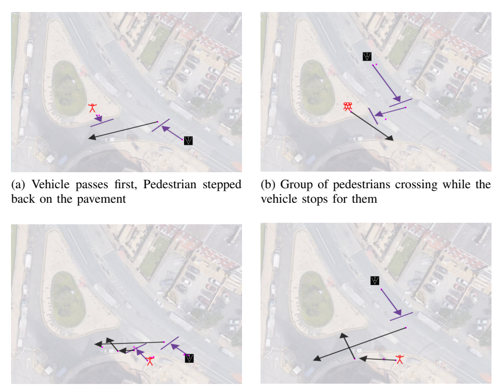
Previous studies have suggested that for autonomous vehicles, some apparently intuitive human communication styles might not be necessary for interactions with pedestrians. [5] showed that facial communication cues such as eye contact do not play a major role in pedestrian crossing behaviors, and that the motion pattern and behavior of vehicles are more important. Human drivers and pedestrians check for eye contact when the vehicle moves in an unexpectedly manner [5]. The field study in [21] showed similar results with an "unmanned" vehicle simulating an autonomous vehicle. This study suggests that similar results could be found with autonomous vehicles. However, [9] showed that pedestrians can use gaze to influence drivers behaviour and make them stop more often at crossings, which has the advantage of increasing the pedestrians' confidence while crossing. Similar results from [20] shows that vehicle movement is sufficient for indicating the intention of drivers and present some motion patterns of road users such as advancing, slowing early and stopping short.

II. Methods

A. Data collection

An ethnographic observation study on pedestrian-vehicle interactions was conducted at an unsignalized intersection near the University of Leeds, UK. After a six-weeks exploration phase, including the observation of 70 pedestrian-vehicle interactions constituting the basis for 15 iterations to design a digital observation protocol, 204 road-crossing interactions were observed in a unified manner including the presence or absence of temporal event features and static descriptor features within each interaction. 62 temporal event types and 12 environmental descriptors were defined, as experimenter hypotheses about what features may be predictive of the interaction winner, as listed and described in Table I. An *environmental descriptor* is a feature of an interaction which does not occur at a single point in time but is present throughout all or most of the interaction, such as agent gender, the weather, and the geometry of the road. An *event* is something which occurs at a single instant in time, such as the pedestrian placing a foot in the road, or the driver giving a signal.

Fig. 1: Intersection where pedestrian-vehicle road-crossing interactions were observed, by observers at locations X and Y. (WGS84: 53.8073, -1.5518)



(a) Vehicle passes first, Pedestrian stepped back on the pavement

(b) Group of pedestrians crossing while the vehicle stops for them

(c) Pedestrian initiating crossing but finally decides to walk further on the left after the vehicle passes

(d) Vehicle stopped due to traffic while the Pedestrian is approaching but the vehicle passed first

Fig. 2: Examples of observed pedestrian-vehicle interactions at the unsignalized intersection

The observers positioned themselves near the intersection, some times in location X and some other times in location Y, as shown in figure 1, and worked together to identify and agree on when a vehicle-pedestrian pair took place in an interaction, one observing the vehicle and driver behaviour and the other observing the pedestrian behaviour. From the very start of the observation, each observer talked out loud about how the observed subject moved, communicated and reacted to the other observer's subject, which allowed the collaborative explication of timely correct behaviour sequences. Once the interaction had ended, both observers filled in the digital observation protocol from start till end, one typing and the other controlling.

Features	Features
0: Approaching Phase: Driver/Vehicle Stopped due to traffic	1: Approaching Phase: Vehicle Used signals Turn Indicator
2: Approaching Phase:Driver/Vehicle approached From left	3: Approaching Phase: Pedestrian Movements Slowed down
4: Approaching Phase: Driver/Vehicle Accelerated	5: Approaching Phase: Driver/Vehicle Turned right
6: Approaching Phase: Pedestrian Head Movements Turned left	7: Approaching Phase: Pedestrian Looked at approaching vehicle
8: Approaching Phase: Pedestrian Stepped on road and stopped	9: Approaching Phase: Driver/Vehicle Passed the pedestrian
10: Crossing Phase: Pedestrian Initiated crossing movement	11: Approaching Phase: Driver/Vehicle Decelerated due to traffic
12: Approaching Phase: Driver/Vehicle Kept pace	13: Approaching Phase: Pedestrian Stopped at the edge of the pavement
14: Crossing Phase: Pedestrian Head Movements Turned right	15: Crossing Phase: Pedestrian Looking at Looked at vehicle
16: Crossing Phase: Pedestrian Movements Other (elaborate in notes)	17: Approaching Phase: Pedestrian Movements Kept pace
18: Approaching Phase: Pedestrian Head Movements Turned right	19: Approaching Phase: Driver/Vehicle Interacting vehicle Van
20: Approaching Phase: Driver/Vehicle Turned left	21: Crossing Phase: Pedestrian Stepped back on pavement
22: Crossing Phase: Pedestrian Head Movements Turned left	23: Approaching Phase: Driver/Vehicle Head Turned in the direction of pedestrian
24: Approaching Phase: Pedestrian Movements Did not Stop	25: Crossing Phase: Pedestrian Slowed down / stopped
26: Approaching Phase: Driver/Vehicle Decelerated for observed pedestrian	27: Crossing Phase: Pedestrian Speeded up
28: Approaching Phase: Driver/Vehicle Interacting vehicle Other (elaborate in Notes)	29: Crossing Phase: Pedestrian Looking at other RUs (elaborate in comments)
30: Approaching Phase: Driver/Vehicle Interacting vehicle Bus / Truck	31: Approaching Phase: Driver/Vehicle Stopped due to other pedestrian
32: Approaching Phase: Driver/Vehicle approached from Multiple	33: Approaching Phase: Driver/Vehicle Used signals Flashed Lights
34: Crossing Phase: Pedestrian Raised hand sidewards	35: Approaching Phase: Driver/Vehicle Decelerated due to other pedestrians
36: Approaching Phase: Pedestrian Looking at other RUs Others (elaborate in notes)	37: Approaching Phase: Pedestrian Looking at other pedestrians entering the road
38: Crossing Phase: Driver/Vehicle Passed the pedestrian'	39: Crossing Phase: Pedestrian Looked at driver
40: Crossing Phase: Driver/Vehicle Stopped for observed pedestrian	41: Crossing Phase: Driver/Vehicle Head Turned in the direction of pedestrian
42: Crossing Phase: Driver/Vehicle Raised hand in front	43: Crossing Phase: Pedestrian Raised hand in front
44: Crossing Phase: Driver/Vehicle Turned right	45: Approaching Phase: Vehicle Stopped for observed pedestrian
46: Crossing Phase: Driver/Vehicle Accelerated	47: Approaching Phase: Pedestrian Speeded up
48: Crossing Phase: Driver/Vehicle Decelerated for observed pedestrian	49: Crossing Phase: Vehicle Waved hand
50: Crossing Phase: Driver/Vehicle Movement Other (elaborate in notes)	51: Crossing Phase: Driver/Vehicle Turned left
52: Approaching Phase: Pedestria Hand Movements Other (elaborate in notes)	53: Approaching Phase: Driver Head Turned right
54: Approaching Phase: Driver/Vehicle Movement Other (elaborate in notes)	55: Approaching Phase: Driver/Vehicle Head Movements Other (elaborate in notes)
56: Approaching Phase: Driver/Vehicle Head Turned left	57: Crossing Phase: Pedestrian Waved Hand
58: Crossing Phase: Pedestrian Hand Movements Other (elaborate in notes)	59: Approaching Phase: Driver/Vehicle Waved hand
60: Crossing Phase: Driver/Vehicle Used signals Turn Indicator	61: Crossing Phase: Pedestrian Looking at other pedestrians entering the road
Driver/Vehicle Interacting Vehicle is Single	Driver/Vehicle Interacting Vehicle Coming From right
Weather: Overcast	Weather: Sunny
Weather: Rainy	Group of Pedestrians
Pedestrian: teenager (13-18y)	Pedestrian: young adult (18-30y)
Pedestrian: midage adult (30-60y)	Pedestrian: older person (60+ years)
Pedestrian's Distraction	Pedestrian: Gender

TABLE I: Selected features for the observation of Pedestrian-Vehicle Interaction

Observation and data collection were conducted in accordance with University of Leeds Ethics and Data Protection regulations.

B. Data Preparation

The winner for each interaction was determined and annotated according to the presence of certain events which are indicative of the outcome, such as *Vehicle passed the Pedestrian* or *Vehicle stopped for observed Pedestrian*. Figure 3 shows an overview of the analysis process.

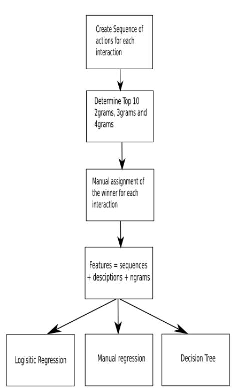


Fig. 3: Diagram for Data Analysis Process

C. Motif Selection

The sequences analysis started by seeking for most common subsequences, some patterns that were occuring more often during interactions. Previous works in bioinformatics such as [6] suggested motif analysis. The top ten *n*-grams (motifs) were extracted from the event sequences for $n \in \{2, 3, 4\}$ as shown in 4, 5 and 6, using BioPython software [4].

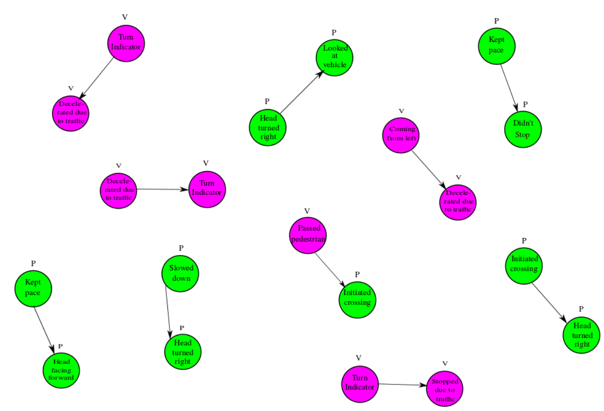


Fig. 4: Best 2-gram of P-V interaction (the pedestrian's actions are in red and vehicle's actions are in green)

D. Logistic Regression

First, Recursive Feature Elimination (RFE) was used to extract the best ten features from the table I. RFE recursively removes less relevant features by building a model on the remaining attributes, leaving only the most relevant features for prediction. The accuracy of this model allows to select the best features that contribute the most to predict the target class. Logistic Regression was then performed on these selected ten features. Logistic regression predicts a binary

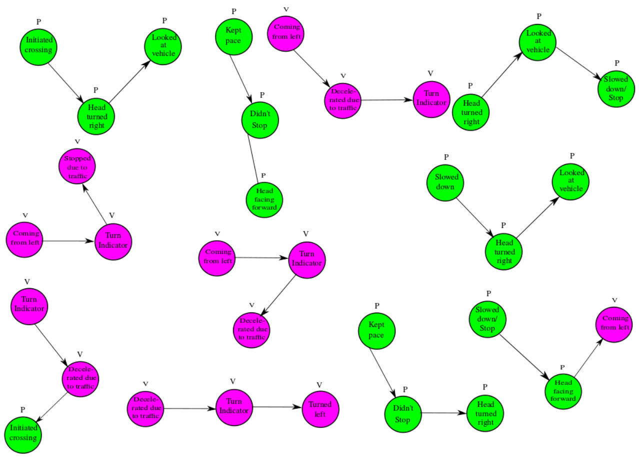


Fig. 5: Best 3-gram of P-V interaction (the pedestrian's actions are in red and vehicle's actions are in green)

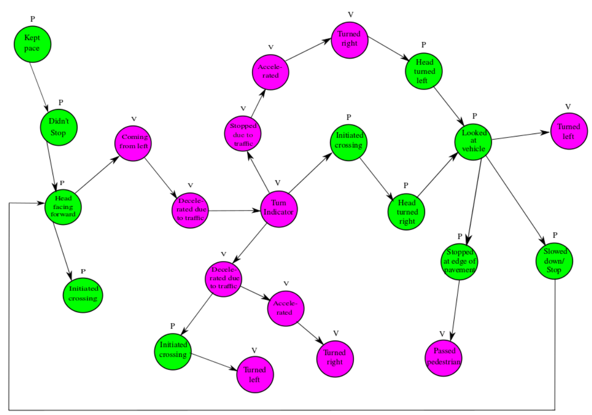


Fig. 6: Best 4-gram of P-V interaction (the pedestrian's actions are in red and vehicle's actions are in green)

dependent (target) variable $y \in 0, 1$ from a vector binary inputs $x_j \in 0, 1$ by optimising parameters of real-valued weights w_j. Weights are set by minimization of the mean squared error between the target y and the sigmoidal model,

$$\hat{y} = \frac{1}{1 + \exp^{-\sum_j w_j x_j}}. \tag{1}$$

The inputs x_i comprised: (1) the presence of absence of each of the 62 actions (a bag-of-words model, ignoring their ordering); (2) the 12 environment descriptors including weather, group size / number of people involved, age and gender of the pedestrian; (3) the presence or absence of the top ten n-grams motifs found previously. In total this gives 104 features. Data were split into a training set composed of 67% of the data and a test set with the remaining 33%. Python scikit-learn library was used for the analysis.

E. Manual Regression

RFE is an automated process which gives no consideration to subjective belief about which features are relevant. This behaviour is desired in cases where it is known that all features are causal of the outcome, but as our model only has a weak concept of time (via the motif subsequencing) the possibility remains that some features may effectively measure effects rather than causes of the target. If this was the case then such features may dominate in RFE and thus remove otherwise interesting causal features. To examine this possibility, we also apply a hypothesize-and-test procedure in which the human experimenters define a set of features X_i, $\{$*Distraction, Age, Gender*$\}$ – which they considered to be both causal and potentially informative, and measure the mutual information about the target distribution Y provided by these individually,

$$H[P(Y); P(Y \mid X_i)] = H[P(Y)] - H[P(Y \mid X_i)], \tag{2}$$

F. Decision Tree Regression

Decision Trees are a greedy relevant feature selection method, alternative to RFE+logit regression above, and which provides some visualisation helpful for human interpretation, and a fast method for real-time systems such as AVs to make decisions based on a few variables. We apply a particular Decision Tree method consisting of finding the best single feature at each step, based on information gain with MSE (Mean Square Error) score for y_i the true target class and $\hat{y}_i$ the predicted value,

$$MSE(y) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

III. Results

A. Prior outcomes

In the absence of any other information, 62% (127 of 204) interactions resulted in the vehicle winning, i.e. passing through the conflict area before the pedestrian. There were no observed collisions between vehicles and pedestrians.

B. Motif selection

Motif selection provides an interesting view of pedestrian-vehicle interactions, by identifying common short sequences of events which tend to occur together. The most common n-grams are shown in figs. 4, 5 and 6, colouring the nodes to denote pedestrian and vehicle events. The most occuring 2-gram motif is (*Pedestrian Head Turned Right, Pedestrian Looks At The Vehicle*), the most common 3-gram motif is (*Pedestrian Initiated crossing movement, Pedestrian Head Turned right, Pedestrian Looked at vehicle*) and the best 4-gram motif is (*Driver/Vehicle Used signals Turn Indicator, Pedestrian Initiated crossing movement, Pedestrian Head Turned right, Pedestrian Looked at vehicle*).

C. Logit regression: automated

The logistic regression results tell us that the ten features extracted from RFE are statistically significant for the model. This is shown in Table II by the absolute z-values higher than 2 and the p-values less than 0.05, which means that the model fits within 95% confidence and that with each feature we are able to predict the winner for each interaction. Especially, the model predicts well on the test set with an accuracy of up to 92%.

Features	Coef	z	p	Winner
1: Approaching Phase: Vehicle Used signals Turn Indicator	-1.2806	-4.672	0.000	Vehicle
7: Approaching Phase: Pedestrian Looked at approaching vehicle	-2.5647	-3.842	0.000	Vehicle
17: Approaching Phase: Pedestrian Movements Kept pace	2.8948	2.490	0.013	Pedestrian
21: Crossing Phase: Pedestrian Stepped back on pavement	-3.9545	-2.323	0.020	Vehicle
25: Crossing Phase: Pedestrian Slowed down / stopped	-2.3499	-3.395	0.001	Vehicle
26: Approaching Phase: Driver/Vehicle Decelerated for observed pedestrian	2.6585	3.660	0.000	Pedestrian
27: Crossing Phase: Pedestrian Speeded up	8.1592	4.118	0.000	Pedestrian
29: Crossing Phase: Pedestrian Looking at other RUs (elaborate in comments)	-3.5777	-2.253	0.024	Pedestrian
41: Crossing Phase: Driver/Vehicle Head Turned in the direction of pedestrian	4.7938	2.034	0.042	Vehicle
Veh. Turn Indicator + Veh. stopped due to traffic	5.0094	4.446	0.000	Vehicle

TABLE II: Logistic Regression Results for the 10 best features

D. Hypothesis testing

Table III shows the informativeness of each of the hypothesised features about the interaction outcome, and the direction of the predictive effect (ie. whether the presence of the feature favours the Pedestrian or the Vehicle to win). The entropy of our outcome classes in the absence of any feature information is 0.956 bits, as the prior probabilities are (62% vehicle wins vs. 38% pedestrian wins). We assume without proof that 1 millibit (0.001 bit) is a significant information gain (as it corresponds roughly, for example, to a change from a 50:50 belief to a 48:52 belief. One full bit of information would correspond to a perfect prediction of the binary class.).

X_i	$H[Y\|X_i](bits)$	$I[Y;X_i](millibits)$	Winner
None	0.956	n/a	Vehicle
Age 60+ years	0.9353	20.8	Pedestrian
Age = 18-30years	0.9496	6.6	Pedestrian
Age = 13-18years	0.9525	3.6	Pedestrian
Distraction	0.9537	2.5	Vehicle
Age = 30-60years	0.9556	0.5	Not significant
Gender	0.9557	0.4	Not significant

TABLE III: Manual Regression Results

The strongest single feature from the set is the presence or absence of elderliness of the pedestrian. (ie. whether they appear to be over 60 years old), which yields 20.8 millibits of information about the outcome (in the present data set, every elderly person wins, though there remains uncertainty about outcome for non-elderly people). Other age classes are also significant, with teenagers (13-18) tending to win, and young working-age adults tending to lose. Distraction then appears as a significant (2.5 millibits) factor (including by their mobile phones, headphones or chatting) in favour of the vehicle winning. Gender appears to be not significant, which was an unexpected finding.

E. Decision trees

We obtain the decision tree shown in figure 7 with the training set. This tree performs on the test set with 58% accuracy.

IV. Discussion

In the absence of other information, the vehicles are more likely (62%) to win an interaction than a pedestrian. This is roughly consistent with the game theory model of [7], which shows how the physically stronger agent is more likely to win, as the asymmetric negative utilities in the (unlikely) event of a collision recurse through the game theory equations to affect yielding in (more likely) non-crash scenarios.

The motif results suggest the way the vehicle and the pedestrian signal their intention of crossing, the former using the turn indicator and the latter looking for clues in the vehicle movement before crossing. Figure 6 is particularly interesting as it shows that the 4-grams are connected to each other, and it highlights these two features *Vehicle using the Turn Indicator - Pedestrian looking at the vehicle* which seems to be the important features for the communication occuring during the interaction. Pedestrian looks for some visual cues and the vehicle needs to understand that the pedestrian has noticed its presence before eventually crosing. Our evidence contributes to the debate over the relevance of human eye-contact vs vehicle position signalling, consistent with [18], [19] in that our pedestrians display their own intention to cross to the vehicle by turning their head and looking at it, as seen in the n-grams. A head-turn is easier for the vehicle to see than eye contact from a fixed head position, which suggests this event may have a dual function both to passively observe the vehicle and to actively signal intent to it. As suggested by the ten automatic extracted features in our data and in accordance with [5] [20], we found that pedestrians then seek for cues in the vehicles *motion* – not in eye contact with or gestures by the driver – then in the surrounding environment such as looking at other road users. Indeed, the data collection observers themselves did not record much information about the driver gestures because they were difficult to see. So eye gaze by the pedestrian is important, but eye contact with the driver or AV is not, as found in [11]. These findings are important for AV design as they suggest that AVs should also be designed to communicate simply via their position on the road (and maybe use of indicator lamps) but maybe not needing artificial face, eye, or gesture substitutes; and that they do need to detect and process pedestrian faces and eyes in order to inform their interactions.

Significant hypothesised features which predict the winners include: age of pedestrian, with elderly people (perhaps more likely to be treated well by drivers for social utility) and teenagers (perhaps more likely to be assertive for their own utility). Equally there are two interesting findings. First, a surveyed "Driver flashes headlights" property does not appear in any significant results, which was hypothesised wrongly by the experimenters to be a common form of communication (meaning "go ahead, I will yield") in the UK. Also gender of pedestrian appears to be not significant, suggesting that both drivers and pedestrians do not discriminate on gender in their behaviours which supports the finding of [15].

The present study considers the predictors almost as a pure bag of words, without regard to the temporal order in which they occur. We added simple motif analysis to allow for a limited notion of temporality in the input, but this was (against our expectation) found to have only a minor predictive effect over the bag-of-words. To extend the model more towards real-world use, more detailed temporal structure could be considered. Suppose that the environment features e_i are all observable at the start of the interaction, and that the temporal event features f_i are revealed to us over time, at times $t(f_i)$. To construct a game-theoretic AV controller to run in real time, it is necessary to make decisions based on partial observations of these features as they are revealed. We must make an action a_t at time t (to yield or not yield to the pedestrian) based on our assessment of $P(\theta_{ped}|\{e_i\}_i, \{f_j : t(f_j) < t\})$. For example, some event features which tend to occur near the end of an interaction may have high predictive value but are of little practical use because the game is almost over at this point. Deciding when to act based on the revealed subsequences will then form an optimal stopping problem and make use of such problems classic solution methods [24]. An early work on the temporality of observed features can be found in [2].

Related to the temporal issue, the reason why the manual hypothesised features of interest have less predictive power than the automatically extracted ones is of interest. Why were humans so poor at guessing the most predictive features? The reason possible is due to a lack of cause and effect separation between the variables. Humans prefer to think of clearly causal variables such as age and gender as predictors of the interaction result, while the automatic methods finds features such as *Pedestrian kept pace* or *Vehicle used turn indicator* which correlate strongly but perhaps as an effect of the more underlying causal variables. Future work could apply causal Bayesian Network modelling [16], to clarify this, and the explicit temporal sub-sequencing above would be likely to result in higher precedence of the human-hypothesis factors in real-time prediction.

The present study only considers the prediction of the interaction outcome – who wins – rather than the inference of the underlying latent assertiveness variabile θ involved in causing the outcome. This is an important first step towards inferring θ, and the winner can to a first approximation be used as a proxy for pedestrian θ. Future work should consider how to make the inference more precisely and also how to infer and separate the effect of the drivers own assertiveness θ_{driver} from the pedestrians. This could be performed, for example, using causal Bayesian network models with EM algorithm to infer the latent variables [22]. As such, the present model does form a complete real-time pedestrian system but it does represent an important research step towards this goal. It has designed and collected a data set of the highest road crossing detail known to us, and shown how predictors can be found. Future work should study

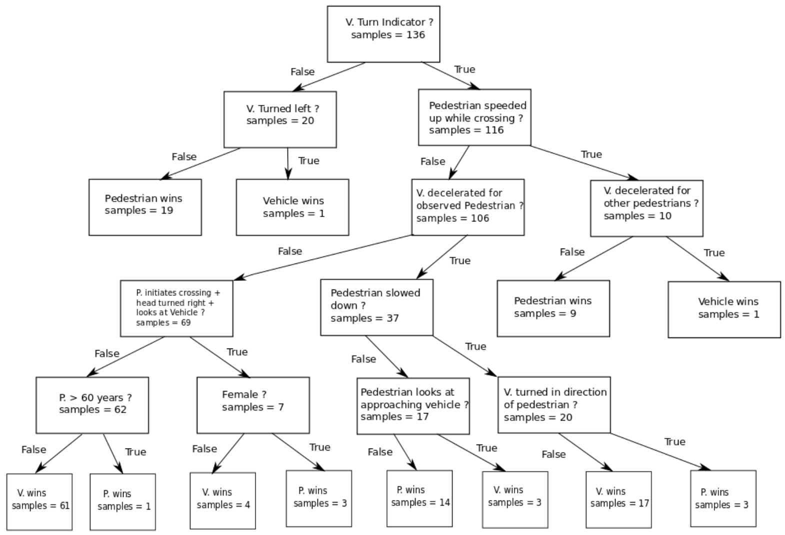


Fig. 7: Decision Tree for Pedestrian-Vehicle Interaction

the best predictors found here and move them into temporal latent variable models along with the game theory model of [7] as the next step towards a real-time AV controller.

References

[1] J. J. Burred. Genetic motif discovery applied to audio analysis. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 361–364, March 2012.

[2] F. Camara, O. Giles, R. Madigan, M. Rothmüller, P. H. Rasmussen, S. A. Vendelbo-Larsen, G. Markkula, Y. M. Lee, L. Garach, N. Merat, and C. W. Fox. Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. In *Proceedings of the 15th International Conference on Intelligent Autonomous Systems workshops*, 2018.

[3] F. Camara, R. Romano, G. Markkula, R. Madigan, N. Merat, and C. W. Fox. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Measuring Behavior 2018*. Manchester Metropolitan University, March 2018.

[4] P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. L. de Hoon. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.

[5] D. Dey and J. Terken. Pedestrian interaction with vehicles: Roles of explicit and implicit communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '17, pages 109–113, New York, NY, USA, 2017. ACM.

[6] P. D'haeseleer. How does dna sequence motif discovery work? *Nature biotechnology*, 24:959–61, 09 2006.

[7] C. W. Fox, F. Camara, G. Markkula, R. Romano, R. Madigan, and N. Merat. When should the chicken cross the road?:game theory for autonomous vehicle - human interactions. In *VEHITS*, January 2018.

[8] A. Gorrini, G. Vizzari, and S. Bandini. Towards modelling pedestrian-vehicle interactions: Empirical study on urban unsignalized intersection. *CoRR*, abs/1610.07892, 2016.

[9] N. Guéguen, S. Meineri, and C. Eyssartier. A pedestrian?s stare and drivers? stopping behavior: A field experiment at the pedestrian crossing. *Safety Science*, 75:87 – 89, 2015.

[10] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada. An open approach to autonomous vehicles. *IEEE Micro*, 35(6):60–68, Nov 2015.

[11] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila. Context-based pedestrian path prediction. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 618–633, Cham, 2014. Springer International Publishing.

[12] M. Liu, Y. Chen, G. Lu, and Y. Wang. Modeling crossing behavior of drivers at unsignalized intersections with consideration of risk perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 45:14 – 26, 2017.

[13] S. Mamidipalli, V. P. Sisiopiku, B. Schroeder, and L. Elefteriadou. A review of analysis techniques and data collection methods for modeling pedestrian crossing behaviors. *Journal of Multidisciplinary Engineering Science and Technology*, Vol. 2, Issue 2, February 2015.

[14] J. MitrovićSimić, V. Basarić, and V. Bogdanović. Pedestrian crossing behaviour at unsignalized crossings. 2017.

[15] E. Papadimitriou, S. Lassarre, and G. Yannis. Pedestrian risk taking while road crossing: A comparison of observed and declared behaviour. *Transportation Research Procedia*, 14:4354 – 4363, 2016. Transport Research Arena TRA2016.

[16] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000.

[17] A. Pillai. Virtual reality based study to analyse pedestrian attitude towards autonomous vehicles. G2 pro gradu, diplomity, 10 2017.

[18] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. Agreeing to cross: How drivers and pedestrians communicate. *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 264–269, 2017.

[19] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles*, 3(1):61–70, March 2018.

[20] M. Risto, C. Emmenegger, E. Vinkhuyzen, M. Cefkin, and J. Hollan. Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. pages 186–192, 11 2017.

[21] D. Rothenbücher, J. Li, D. Sirkin, B. Mok, and W. Ju. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. pages 795–802, 08 2016.

[22] S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11(2):305–345, 1999.

[23] F. Schneemann and I. Gohl. Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pages 38–43, June 2016.

[24] A. N. Shiryaev. Optimal stopping rules. In *International Encyclopedia of Statistical Science*, 2011.

[25] D. Sun, S. Ukkusuri, R Benekohal, and S. Travis Waller. Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks. 51, 12 2002.